Online tool for Massive Observational Studies (OMOS) in Drug Repositioning and Cancer Prevention: A Technical Overview

Phung-Anh (Alex) Nguyen, Ph.D.[1], Usman Iqbal, PharmD, MBA[1], Chih-Wei Huang, MS[1], Shen-Hsien Lin, MS[1], Wen-Shan Jian, Ph.D.[3], Yu-Chuan (Jack) Li, M.D., Ph.D.[1,2,*]
[1]College of Medical Science & Technology, Taipei Medical University, Taiwan; [2]Department of Dermatology, Wan-Fang Hospital, Taiwan; [3]School of Health Care Administration, Taipei Medical University, Taiwan; * Corresponding author

Abstract

Health observational data also known as big-data is leveraging every day. To utilize this is an effective way for generating health outcomes of interest, especially in long-term used drugs and cancer risk. Methodology designed with experts help based upon existing statistical methods for pharmacoepidemiology studies. Model of case-controls helped to generate results automatically by using conditional logistic regressions, in order to do mass production studies, saving time, cost effective and do not require professionals.

Introduction

Pre-marketing drug evaluation processes are still largely based on formal, extremely expensive randomized clinical trials and manual information collection processes that cover a relatively small sample of patients.[1] Moreover, the rapid change in health information technology system had dramatically increased health data accumulated.[2] It has become an important material that provides an extraordinary opportunity to observe the emergence of new knowledge and its influence, particularly in drug repositioning and cancer prevention.[3] Upon this beneficial, we aimed to develop an online informatics tool in order to evaluate the risk of drugs for cancer by utilizing medical big data. It could produce the massive studies, reduce cost, time verses traditional trials, and help to improve drug safety, quality in health care.

Methods

Data Source: We use the Taiwan's National Health Insurance Database that has provided a huge data which covered all health information including characteristics and all drug information i.e. prescriptions, etc. of 23 million Taiwanese population.
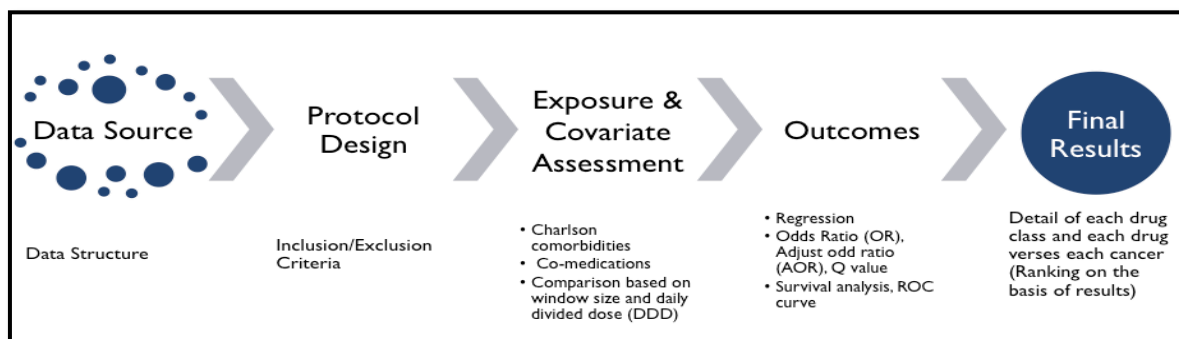

Figure 1. The processing of OMOS system

System overview: The processing of OMOS system is consists of five steps, which are shown in Figure 1.

Front-end development: Web-based interface was developed by using PHP package and Javascript. In addition, we included the guidelines of evidence based medicine (EBM) level 3 for observational study such as cohort, case-control, and/or case serial self-control in order to support users interact with system.[4]

Back-end development: A package of Apache, MySQL & PHP was used to build the serve-side of the system. We integrated the Elasticsearch API[5] to our system in order to search and analyze data immediately. The example of data transform to person-level from Taiwan NHI database is shown in Box 1. After then, we also integrated the analytics package (i.e. R package) to perform the statistical analysis to a given study.
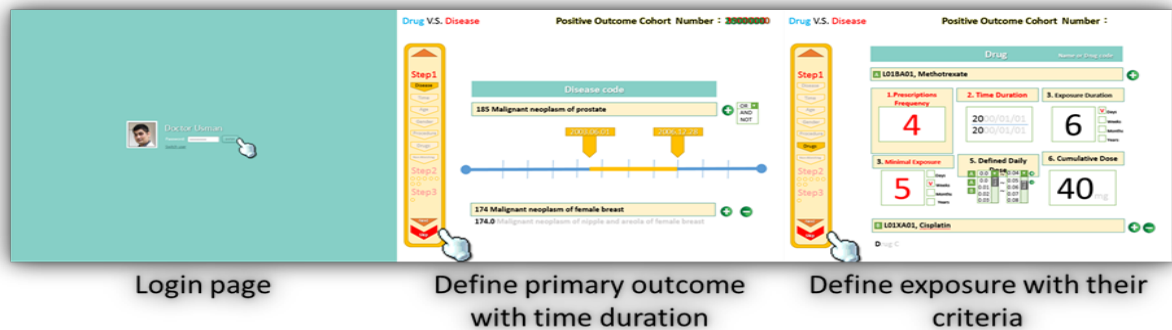
Results

We successfully developed and built the informatics tool, which used large observational health data to study the effects of medical interventions and predict health outcomes, especially for long-term use drug verse cancer risk. Our system has been design to satisfy few requirements such as mass production studies, save time, cost effectiveness, and do not require professionals. The UI of our system is shown in Figure 2.

Box 1. An example of data transform from Taiwan's NHI claim database

```
"cd": { //Records of diseases
"properties": {
          "birthday_year": {"type": "integer"},
          "gender": {"type": "string"},
          "pid": {"type": "string"},
          "sn": {"type": "long"},
          "subtotal": {"type": "nested",
                    "properties": {
                    "count":{"type": "integer"},
                    "first_days":{"type": "integer"},
                    "icd9": {"type": "string"},
                    "last_days": {"type": "integer"}
                    }
          }
    }
}
```

```
"oo_drug": { //Records of drugs use
"properties": {
          "birthday_year": {"type": "string"},
          "gender": {"type": "string"},
          "oo_drug": {"type": "nested",
                    "properties": {
                    "atc_code": {"type": "string"},
                    "days": {"type": "integer"},
                    "drug_day": {"type": "integer"},
                    "drug_no": {"type": "string"},
                    "drug_use": {"type": "integer"}
                    }
          },
          "pid": {"type": "string"},
          "sn": {"type": "long"}
    }
}
```

Figure 2. An example of UIs in the system

Login page | Define primary outcome with time duration | Define exposure with their criteria

Discussions and Conclusions

Developing a model which is capable for long term drug use and cancer risk on a societal scale is a big challenge that is approachable, achievable, and has implications towards those developing and/or using medications. Such research model would also provide an excellent test bed for solving the technological, informatics, and organizational issues towards other broad domains of drug evaluation mass production studies by utilizing large-scale databases.

References

1. Hiatt WR. Observational Studies of Drug Safety — Aprotinin and the Absence of Transparency. *New England Journal of Medicine.* 2006;355(21):2171-2173.
2. Blumenthal D, Tavenner M. The "Meaningful Use" Regulation for Electronic Health Records. *New England Journal of Medicine.* 2010;363(6):501-504.
3. Benson K, Hartz AJ. A Comparison of Observational Studies and Randomized, Controlled Trials. *New England Journal of Medicine.* 2000;342(25):1878-1886.
4. Doi SAR. *Understanding evidence in health care: Using clinical epidemiology.* South Yarra, VIC, Australia: Palgrave Macmillan; 2012.
5. BV E. Elasticsearch. https://www.elastic.co/. Accessed Sep 07, 2015.

The committee had a few comments that we'd like to see addressed in a revised submission for acceptance:

Question 1. There is no mention of OHDSI or the OMOP common data model. Does the application run against the OMOP CDM currently? If not, would it be possible to revise it to make that possible? To what extent do you envision your framework being applied across the OHDSI network such that the aggregate summary results from multiple institutions could be shared and compared alongside your results from the Taiwan NHI database?

Answer: Thanks for your comments. We developed the OMOS tool, which it did not use either OMOP common data model from OHDSI or Taiwan's National Health Insurance (NHI) data structures. As we mentioned in method section, we transformed our NHI data structures to the typed JSON by following the Elasticsearch API. It could help the system performed 3-fold faster than traditional NHI data structures as using mySQL and/or MS SQL server to store and search the data needed (ie. 2.2mins vs. 7mins per query working in the same database with 200 million records for the Elasticsearch and MS SQL server respectively). It would be possible to revise in order to working with different types of data in CDM.

Therefore, if the framework could be applied across the OHDSI network, it could help summary results from multiple institutions by using same UIs (i.e. same format of tables and figures), saving time, cost, and do not require professional as using the same protocol.

Question 2. There is no link to where the source code can be found. Do you intend for your tool to be an open-source application that others within the OHDSI community could use within their own environments and also contribute back to?

Answer: Thanks for your comments. We would like to introduce the technical overview of the OMOS tool in this submission as we have just finished the first version with limited features. Other institutions in OHDSI community are welcome to use and contribute to develop it as we also aim to make it to be an open-source application.

Question 3. There are no analysis results provided in the abstract. Could you please clarify the analytical use case? It is ok for the poster to focus on the tool and framework, without having summary results from Taiwan presented, but it would be useful to better understand what type of expected outputs the system would create, and how you would anticipate a researcher interpretting those findings.

Answer: Thanks for your important question. Mentioned in method section (Front-end development), we included the guidelines of evidence based medicine (EBM) level 3 for observational study such as cohort, case-control, and/or case serial self-control in order to support users interact with system. In the first version of OMOS, we used case-control as a main guideline, particularly in studies with drug use and risk of cancer. The conditional logistic regression were adjusted for potential confounders and used to investigate the association between exposure to the different drugs and risk for cancer. Moreover in future work of cohort, we would consider proportional cox regression models with the time (in days) as the time scales, which use to calculate hazard ratio with 95% CI, etc.